DOCUMENT RESUME

ED 085 651 CG 008 607

TITLE	Some Comments on "An Analysis-of-Variance Model for the Intrasubject Replication Design."
INSTITUTION	Stanford Univ., Calif. Stanford Center for Research and Development in Teaching.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	RDM-111
PUB DATE	Aug 73
CONTRACT	NE-C-00-3-0061
NOTE	11p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Analysis of Variance; *Classroom Research; Data Analysis; *Educational Improvement; Intervention; Models; Research and Development Centers; *Research Design; Research Methodology; *Research Reviews (Publications)

ABSTRACT

This memorandum, part of a series, critically examines the fixed-effects analysis of variance procedure used in an intensive classroom study. The procedure, used by Gentile, Roden and Lein (Journal of Applied Behavioral Analysis, 1972, 5) for intensive studies of single subjects, is found inappropriate because the basic assumptions of an analysis of variance model are typically violated when continuous data on a single subject is gathered over time. Two other proposed analyses of variance models for single subjects are also considered and found unsuitable. Time series analysis, which takes serial correlation effects into account, and a median-based method are recommended as alternatives to analysis of variance designs. (Author/LRP)

CG

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN-
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRE-
SENT OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

SC
RD
T

Stanford Center for Research and Development in Teaching

SCHOOL OF EDUCATION STANFORD UNIVERSITY

FILMED FROM BEST AVAILABLE COPY

ED 085651

STANFORD CENTER
FOR RESEARCH AND DEVELOPMENT
IN TEACHING

Research and Development Memorandum No. 111

SOME COMMENTS ON "AN ANALYSIS-OF-VARIANCE MODEL FOR THE INTRASUBJECT REPLICATION DESIGN"

Carl E. Thoresen and Janet D. Elashoff

School of Education
Stanford University
Stanford, California

August 1973

Published by the Stanford Center for Research and Development in Teaching, supported in part as a research and development center by funds from the National Institute of Education, U. S. Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the National Institute of Education. (Contract No. NE-C-00-3-0061.)

Introductory Statement

The Center's mission is to improve teaching in American schools. Too many teachers still employ a didactic style aimed at filling passive students with facts. The teacher's environment often prevents him from changing his style, and may indeed drive him out of the profession. And the children of the poor typically suffer from the worst teaching.

The Center uses the resources of the behavioral sciences in pursuing its objectives. Drawing primarily upon psychology and sociology, but also upon other behavioral science disciplines, the Center has formulated programs of research, development, demonstration, and dissemination in three areas. Program 1, Teaching Effectiveness, is now developing a Model Teacher Training System that can be used to train both beginning and experienced teachers in effective teaching skills. Program 2, The Environment for Teaching, is developing models of school organization and ways of evaluating teachers that will encourage teachers to become more professional and more committed. Program 3, Teaching Students from Low-Income Areas, is developing materials and procedures for motivating both students and teachers in low-income schools.

The intensive experimental research design (N = 1), in which the effects of various interventions on a single subject are studied over time, offers a powerful strategy for understanding teaching and learning processes. A variety of intensive designs as well as analysis methods are available. The present memorandum critically examines a particular data analysis procedure used in an intensive classroom study.

Abstract

The fixed-effects ANOVA procedure used by Gentile, Roden, and Klein (Journal of Applied Behavior Analysis, 1972, 5) for intensive studies of single subjects is found inappropriate. Two other proposed ANOVA models for single subjects are also considered and found unsuitable. Time series analysis, taking into account serial correlation effects, and a median-based method are recommended as alternatives to ANOVA designs.

SOME COMMENTS ON "AN ANALYSIS-OF-VARIANCE MODEL FOR THE INTRASUBJECT REPLICATION DESIGN"

Carl E. Thoresen and Janet D. Elashoff

Gentile, Roden, and Klein (1972) have identified an important problem in data analysis for the applied researcher. Often the data from intensive studies of single subjects over time fail to provide clear-cut evidence of significant behavior change. Reliance on visual inspection as a basis for decision making is often invalid. White (1971), for example, demonstrated that individuals vary widely in their interpretation of data based on visual inspection--even to the point that some interpreted a trend as accelerating while others judged the same trend to be decelerating. Many years ago, Huff (1954) showed how easily the eye could be misled by graphs and charts which distort the data. Obviously, then, there is a need for applied researchers to employ statistical techniques in drawing conclusions about what happens to data within and between phases.

Gentile, Roden, and Klein (1972), acknowledging this problem, proposed a simple analysis of variance approach to studying changes in the subject over time. Their article reports a classroom study involving two students. There are several serious problems, however, in using a standard analysis of variance with such repeated measures data. As Hartmann (1973) points out, the basic assumptions of an analysis of

---

This paper was written in response to an article by Gentile, Roden, and Klein (1972) in the Journal of Applied Behavior Analysis and will be published in the same journal along with other comments.

variance model are typically violated when continuous data on the same subject are gathered over time. These assumptions include (a) a normal distribution of error components, (b) homogeneity of variance of error components, and (c) the independence of error components. Hartmann appropriately points out that the last assumption, that of independence, is an assumption violated with fatal consequences. Serial correlation in the data tends to inflate the degrees of freedom involved and also lowers the variability within phases, thereby yielding a positively biased F ratio.

Hartmann also raises a crucial question about the marked limitations of relying on a mean value and deviations around a mean within a phase, rather than looking at the performance trend within a phase. Indeed, the major advantage of intensive designs is that they avoid the "static" reliance on a mean performance and allow the investigator to examine change within a phase over time (Sidman, 1960; Thoresen, in press). Applied researchers are well aware of the fact that two phases can have identical mean values, yet the slope or trend of the data in one phase can be sharply accelerating while that of a second phase is dramatically decelerating. Hence, reliance on analytic models that only consider variability around a mean performance ignore what might be called the "dynamic" aspects of intensive designs. While Hartmann (1973) has identified major problems with the Gentile-Roden-Klein strategy, some additional observations are worth noting.

Gentile, Roden, and Klein err in assuming that the dependent variable, number of on-task behaviors, has a binomial distribution.

It is most unlikely, given the description of the experiment, that two successive observations of on- or off-task behavior are independent. In such an experiment it would be preferable to use relative frequencies of on-task behavior by observation period or by task as the unit of analysis. The problem of non-independence from observation to observation or from treatment to treatment is not resolved by the combining of phases (A_1+A_2, B_1+B_2). Such a combination does not deal with the important problem of serial correlation effects within each phase. Any positive correlation of observations within a phase yields a positively biased F ratio. In addition, we can also expect the "true probability" of on-task behavior to change across time during a phase. Such a change violates the assumption necessary for the binomial; i.e., each trial has the same probability of success. In fact, Gentile, Roden, and Klein present evidence that the "true probability" of success differs between phases for the same treatment. For example, the proportion of on-task behavior for James (one of the subjects) when compared for the first phase (A_1) and the fourth phase (A_2) yields a χ^2 value (4.39) significant at the .05 level. Hence, pooling the scores for A_1 and A_2 definitely leads to a violation of the binomial assumption.

Interestingly, the analysis of variance model may not even be appropriate for the idealized coin-tossing experiment that Gentile, Roden, and Klein describe. If the coin iteself is not allowed to adapt to the surrounding temperature before beginning each phase, and if the warming up or cooling down phase is included within the data for a particular phase, the basic assumptions of the analysis of variance model are violated.

Other points merit comment. First, there is no logical basis for letting the number of observation periods vary as widely in each phase as Gentile, Roden, and Klein (Table 1, p. 195) allow. The range is approximately five-fold, from 210 observations in one phase to almost 1,000 observations in another phase. Second, conclusions about the effects of treatments for James and Lynn, the two subjects, hold only when these subjects are considered as a fixed effect. If these two were considered as a random sample of subjects with generalizations to be made to a population of similar subjects, the F test for treatments would have been insignificant (3.48, df=2,2).

Finally, it should be noted that the proposed "t-test analysis," where only two treatments and one subject are involved, is identical to the analysis of variance.

Hartmann's Alternative

Hartmann offers an idealized model (his Fig. 1) for data involved in a reversal design. He appropriately points out that before using an ANOVA model one must first test for the assumption of independence, i.e., serial correlation. In addition, there must also be a sufficient number of data points that are "stable" in each of the four treatment conditions. Some problems exist, however, with the Hartmann model. First, failure to find a significant serial correlation of Lag 1 (that is, is Observation No. 1 independent of Observation No. 2, No. 2 independent of No. 3, and so on?) does not guarantee independence. There may be a systematic bias within a phase represented by a Lag 5 relationship so that, for example, a teacher's behavior on Mondays and Fridays

is highly correlated while Monday-Tuesday and Tuesday-Wednesday comparisons do not show significant correlations. Second, tests of correlation coefficients are not very powerful unless sample sizes are large.

Hartmann's suggestion that the analysis incorporate only the "last n data points in each condition obtained during asymptotic responding," although a plausible suggestion, may present difficulties in many real situations. Typically, the data pattern within a phase is more likely to be accelerating, decelerating, or curvilinear. Thus, even if the regression of time on the dependent variable has a zero slope within a phase, it may not correspond to the last few data points within a phase. In practice, it is not easy to identify an interval when data are "stable."

The ANOVA model suggested by Shine and Bower (1971) also offers little solace to the applied researcher. These authors in effect propose a two-way fixed-effects analysis of variance model with one observation per cell. Its appropriateness is limited to a special case where responses are in no way sequentially dependent within treatments, although there may be restricted types of correlation patterns between treatments. Applied researchers seldom deal with behavior that is completely independent from observation to observation.

Alternatives to ANOVA Designs

A preferred strategy to ANOVA is based on various time series analyses (e.g., Gottman, McFall, & Barnett, 1969). The best solution to "noisy" data about which the researcher wishes to make some inferences may be found in analysis techniques that systematically take into account serial correlation effects. Glass, Willson, and Gottman (1973) offer an excellent methodological discussion of various intensive or time

series designs, especially concerning the problems of confounding factors with repeated measures. These authors, building on earlier efforts (e.g., Box & Tiao, 1965), offer what is called an "integrated moving average" method. This procedure allows the researcher to make probability statements about changes in level and slope between treatment phases. A recent example of this procedure is reported by Gottman and McFall (1972) in a study of self-monitoring effects in a high school classroom. An alternative method, based on the use of median-derived slopes to describe progress within and between phases, has been suggested by White (1972). White has used this method with a large number of classroom intervention studies to examine changes in level and slope between phases, such as baseline and intervention. The advantages of this median-based method over a standard regression analysis strategy are currently being examined (see White, 1971). Some questions exist, for example, about whether the median slope procedure adequately deals with the effects of serial dependence.

A thorough discussion of these procedures and others is beyond the scope of this brief memorandum. However, the applied researcher should know that some appropriate methods for analyzing intensive experiments are available in the literature. It is to be hoped that the next few years will see an expansion of efforts to develop appropriate statistical methodologies for intensive research designs.

References

Box, G. E. T., & Tiao, G. C. A change in level of non-stationary time series. Biometrika, 1965, 52, 181-192.

Gentile, J. R., Roden, A. H., & Klein, R. D. An analysis-of-variance model for the intrasubject replication design. Journal of Applied Behavior Analysis, 1972, 5, 193-198.

Glass, G. V., Willson, V. L., & Gottman, J. M. Time series analysis in the behavioral sciences. Boulder, Colo.: Laboratory of Education Research, University of Colorado, 1973.

Gottman, J. M., McFall, R. M., & Barnett, J. T. Design and analyses of research using Time Series. Psychological Bulletin, 1969, 72, 299-306.

Gottman, J. M., & McFall, R. M. Self-monitoring effects in a program for potential high school dropouts: A Time series analysis. Journal of Consulting and Clinical Psychology, 1972, 39, 273-281.

Hartmann, D. P. Forcing square pegs into round holes: Some comments on "An analysis-of-variance model for the intrasubject replication design." Journal of Applied Behavior Analysis, forthcoming (1973). In response to Gentile, Roden, and Klein, 1972.

Huff, D. How to lie with statistics. New York: Norton, 1954.

Shine, L. C. II, & Bower, S. M. A one-way analysis of variance for single-subject designs. Educational and Psychological Measurement, 1971, 31, 105-113.

Sidman, M. Tactics of scientific research. New York: Free Press, 1960.

Thoresen, C. E. The intensive design: An intimate approach to research. The Counseling Psychologist, in press.

White, O. R. Pragmatic approaches to the progress in the single case. Unpublished doctoral dissertation, University of Oregon, 1971.

White, O. R. A manual for the calculation and use of the median slope: A technique of progress estimation and prediction in the single case. Working Paper No. 16. Regional Resource Center for Handicapped Children, University of Oregon, Eugene, Oregon, 1972.